

Natural language processing in biomedicine

- Adamic LA, Wilkinson D, Huberman BA, Adar E. 2002. A literature based method for identifying gene-disease connections. Proc. IEEE Computer Society Bioinformatics Conference, pp. 109-17.
- Amaral MB, Roberts A, Rector AL. 2000. NLP techniques associated with the OpenGALENontology for semi-automatic textual extraction of medical knowledge: abstracting and mapping equivalent linguistic and logistic constructs. Proc AMIA Symp, pp. 76-80.
- Andrade MA, Bork P. Automated extraction of information in molecular biology. FEBS Letters 2000;476:12-7.
- Andrade Miguel A., and Alfonso Valencia. 1998. Automatic annotation for biological sequences by extraction of keywords from MEDLINE abstracts: Development of a prototype system. Bioinformatics 14(7):600-7.
- Aronson AR. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp, pp. 17-21.
- Aronson AR, Bodenreider O, Chang HF, et al. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001;:17-21.
- Baclawski K, Cigna J, Kokar MM, Mager P, Indurkhy B. Knowledge representation and indexing using the unified medical language system. Pac Symp Biocomput., 2000:493-504.
- Bangalore A, Thorn KE, Tilley C, Peters L. 2003. The UMLS Knowledge Source Server: An object model for delivering UMLS data. Proc AMIA Symp.
- Baud RH, Lovis C, Rassinoux AM, Scherrer JR. 1998. Alternative ways for knowledge collection, indexing and robust language retrieval. Methods Inf Med 37(4-5):315-26.
- Blaschke, C.; Andrade, M.A.; Ouzounis, C.; and Valencia, A. (1999). "Automatic extraction of biological information from scientific text: protein-protein interactions." In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, 60-70.
- Bodenreider O, Pakhovov SV. 2003. Exploring adjectival modification in biomedical terms across two genres. In Proc. Biomedical Workshop at ACL
- de Bruijn, B., and Martin, J. (2002). "Getting to the (c)ore of knowledge: mining biomedical literature." *International Journal of Medical Informatics*, 67, 7-18.
- Christensen L, Haug PJ, Fiszman M. MPLUS: A probabilistic medical language understanding system. 2002. Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, pp. 29-36. Association for Computational Linguistics.
- Craven, Mark, and Johan Kumlien. 1999. Constructing biological knowledge bases by extracting information from text sources. ISMB.
- Draselia N, Yruev A, Egorov S, et al. 2004. Extracting human protein interactions from MEDLINE using a full-sentence parser. Bioinformatics 20(5):604-11.

- Elkins JS, Friedman C, Boden-Albala B, Sacco RL, Hripcsak G. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. *Comput Biomed Res* 2000 Feb;33(1):1-10.
- Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. 2000. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 7(6):593-604.
- Fiszman M, Haug PJ. 2000. Using medical language processing to support real-time evaluation of pneumonia guidelines. *Proc AMIA Symp*, pp. 235-9.
- Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. 1994. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1(2):161-74.
- Friedman C, Hripcsak G. 1999. Natural language processing and its future in medicine. *Acad Med* 74(8):890-5.
- Friedman C, Kra P, Yu H, Krauthammer M, Rzhetsky A. 2001. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics* 17 Suppl 1:S74-82.
- Fukuda K, Tsunoda T, Tamura A, Takagi T. Toward information extraction: Identifying protein names from biological papers. *Pac. Symp. Biocomput.*, 1998, 707-18.
- Gaizauskas R, Demetriou G, Artymiuk PJ, Willett P. 2003. Protein structures and information extraction from biological texts: The PASTA System. *Bioinformatics* 19(1):135-43.
- Grishman R, Huttunen S, Yangarber R. 2002. Information extraction for enhanced access to disease outbreak reports. *J Biomed Inform* 35(4):236-46.
- Gover, Clair; Mirella Lapata; Alex Lascarides. 20202. A comparison of parsing technologies for the biomedical domain. *Natural Language Engineering* 1(1):1-38.
- Hahn U. 1989. Making understanders out of parsers: Semantically driven parsing as a key concept for realistic text understanding applications. *International Journal of Intelligent Systems* 4(3):345-93.
- Hahn U, Romacker M, Schulz S. 1999. How knowledge drives understanding--matching medical ontologies with the needs of medical language processing. *Artif Intell Med* 15(1):25-51.
- Hahn U, Romacker M, Schulz S. 2000. MEDSYNDIKATE--design considerations for an ontology-based medical text understanding system. *Proc AMIA Symp*, pp. 330-4.
- Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff S. 1994. A natural language understanding system combining syntactic and semantic techniques. *Proc Annu Symp Comput Appl Med Care*, pp. 247-51.
- Hatzivassiloglou, Vasileios; Pablo A. Duboué; and Andrey Rzhetsky. 2001. Disambiguating proteins, genes, and RNA in text: a machine learning approach. *Bioinformatics* 17(suppl. 1): S97-S106.
- Hishiki T, Collier N, Nobata C, et al. Developing NLP Tools for Genome Informatics: An Information Extraction Perspective. *Genome Informatics Series : Proceedings of the Workshop on Genome Informatics*, 1998;9:81-90.
- Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. 1995. Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 122(9):681-8.

- Humphrey S. 1999. Automatic indexing of documents from journal descriptors: A preliminary investigation. *JASIS* 50(8):661-674.
- Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. 1998. The Unified Medical language System: An informatics research collaboration. *J Am Med Inform Assoc* 5(1):1-11.
- Jain NL Friedman C. 1997. Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. *Proc AMIA Symp*, pp. 829-33.
- Johnson SB, Aguirre A, Peng P, Cimino J. 1993. Interpreting natural language queries using the UMLS. *Proc Annu Symp Comput Appl Med Care*, pp. 294-8.
- Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. *Infect Control Hosp Epidemiol* 1998 Feb;19(2):94-100.
- Krauthammer, Michael; Andrey Rzhetsky; Pavel Morozov; and Carol Friedman. 2000. Using BLAST for identifying gene and protein names in journal articles. *Gene* 259:245-52.
- Kumar A, Smith B. The Unified Medical Language System and the Gene Ontology: Some Critical Reflections. A. Günter, R. Kruse and B. Neumann (eds.), *KI 2003:Advances in Artificial Intelligence* (Lecture Notes in Artificial Intelligence 2821), Berlin: Springer, 2003, 135–148.
- Leroy G, Chen H, Filling preposition-based templates to capture information from medical abstracts. *Pac. Symp. Biocomput.*, 2002:350-361.
- Leroy G, Chen H, Martinez JD. 2003. A shallow parser based on closed-class words to capture relations in biomedical text. *J Biomed Inform.* 36(3):145-58.
- Liu H, Aronson AR, Friedman C. 2002. A Study of Abbreviations in MEDLINE Abstracts. *Proc AMIA Symp*, pp. 464-9.
- Liu H, Johnson SB, Friedman C. Automatic resolution of ambiguous terms based on machine learning and conceptual relations in the UMLS. *J Am Med Inform Assoc.* 2002;9(6):621-36.
- Liu H, Lussier YA, Friedman C. Disambiguating ambiguous biomedical terms in biomedical narrative text: an unsupervised method. *J Biomed Inform.* 2001 Aug;34(4):249-61.
- Marcotte EM, Xenarios I, Eisenberg D. 2001. Mining the literature for protein-protein interactions. *Bioinformatics* 17(4):359-63.
- McCray AT. 1993. Representing biomedical knowledge in the UMLS Semantic Network. *High-Performance Medical Libraries: Advances in Information Management for the Virtual Era*. Meckler Publishing, pp. 45-55.
- McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A and Srinivasan S. 1993. UMLS knowledge for biomedical language processing. *Bulletin of the Medical Library Association* 81:184-194.
- McCray AT, Srinivasan S, Browne AC. 1994. Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care*, pp. 235-9.
- McCray AT, Burgun A, Bodenreider O. 2001. Aggregating UMLS semantic types for reducing conceptual complexity. *Medinfo* 10(Pt 1):216-20.

- Mendonca EA, Johnson SB, Seol YH, Cimino JJ. Analyzing the semantics of patient data to rank records of literature retrieval. Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain of the Association of Computational Linguistics. July 2002,:69-75.
- Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. *J Am Med Inform Assoc*. 2001 Jan-Feb;8(1):80-91.
- Pakhomov SV, Ruggieri A, Chute CG. 2002. Maximum entropy modeling for mining patient medication status from free text. *Proc AMIA Symp*:587-91.
- Park JC, Kim HS, Kim JJ. Bidirectional incremental parsing for automatic pathway identification with combinatory categorial grammar. *Pac. Symp. Biocomput.*, 2001:396-407.
- Proux D, Rechenmann F, Julliard L, Pillet V V, Jacq B. Detecting Gene Symbols and Names in Biological Texts: A First Step toward Pertinent Information Extraction. *Genome Inform Ser Workshop Genome Inform*. 1998;9:72-80.
- Pustejovsky J, Castano J, Zhang J, Kotecki M, Cochran B. 2002. Robust relational parsing over biomedical literature: extracting inhibit relations. *Pac Symp Biocomput*, pp. 362-73.
- Rector AL, Nowlan WA. The GALEN project. *Comput Methods Programs Biomed* 1994 Oct;45(1-2):75-8.
- Rosario B, Hearst M, Fillmore C. 2002. The descent of hierarchy, and selection in relational semantics. Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, pp. 247-54. Association for Computational Linguistics.
- Rassinoux AM, Wagner JC, Lovis C, Baud RH, Rector A, Scherrer JR. 1995. Analysis of medical texts based on a sound medical model. *Proc Annu Symp Comput Appl Med Care*, pp. 27- 31.
- Rosario B, Hearst M, Fillmore C. The descent of hierarchy, and selection in relational semantics. Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, Association for Computational Linguistics July 2002,:247-54.
- Rosse C, Mejino JL. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003 Dec;36(6):478-500.
- Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. *J Am Med Inform Assoc* 1994 Mar-Apr;1(2):142-60.
- Savova GK, Harris M, Johnson T, Pakhomov SV, Chute CG. A Data-driven Approach for Extracting "the Most Specific Term" for Ontology Development. *Proc AMIA Symp*. 2003;:579-83.
- Sekimizu T, Park HS, Tsujii J. Identifying the Interaction between Genes and Gene Product Based on Frequently Seen Verbs in Medline Abstracts. *Genome Informatics Series: Proceedings of the Workshop on Genome Informatics*, 1998;9:62-71.
- Solbrig HR, Elkin PL, Ogren PV, Chute CG. A formal approach to integrating synonyms with a reference terminology. *Proc AMIA Symp*. 2000;:814-8.
- Spyns P. Natural language processing in medicine: an overview. *Methods Inf Med* 1996 Dec;35(4-5):285-301.

- Stephens M, Palakal M, Mukhopadhyay S, Raje R, and Mostafa J. Detecting Gene Relations from MEDLINE Abstracts. Pac. Symp. Biocomput., 2001:483-96.
- Stephens, M.; Palakal, M.; Mukhopadhyay, S.; and Raje, R. (2001). "Detecting gene relations from Medline abstracts." In *Proceedings of the Sixth Pacific Symposium on Biocomputing*, 6, 483-96.
- Taira RK, Soderland SG. 1999. A statistical natural language processor for medical reports. Proc AMIA Symp. 1999;:970-4.
- Tanabe L, Wilbur WJ. 2002. Tagging gene and protein names in biomedical text. Bioinformatics. 18(8):1124-32.
- Tao, Y-C., and Leibel, R.L. (2002). "Identifying relationships among human genes by systematic analysis of biological literature." *BMC Bioinformatics*, 3,16-25.
- Temkin, J. M., and Gilder, M. R. (2003). "Extraction of protein interaction information from unstructured text using a context-free grammar." *Bioinformatics*, 19(16), 2046-53.
- Thomas J, Milward D, Ouzounis C, Pulman S, Carroll M. Automatic extraction of protein interactions from scientific abstracts. Pac Symp. Biocomp., 2000: 538-49.
- Wren JD, Garner HR. 2002. Heuristics for identification of acronym-definition patterns within text: towards an automated construction of comprehensive acronym-definition dictionaries. Methods Inf Med 41(5):426-34.
- Yakushiji A, Tateisi Y, Miyao Y, Tsujii J. Event extraction from biomedical papers using a full parser. Pac. Symp. Biocomput., 2001:408-419. Pac Symp. Biocomp., 386-97.
- Yandell, M.D., and Majoros, W.H. (2002) "Genomics and natural language processing." *Nature Reviews Genetics*, 3, 601-610.
- Yu H, Hripcsak G, Friedman C. 2002. Mapping abbreviations to full forms in electronic articles. J Am Med Inform Assoc 9(3):262-72.
- Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. 1995. A multi-lingual architecture for building a normalised conceptual representation from medical language. Proc Annu Symp Comput Appl Med Care, pp. 357-61.

Rindflesch References (by date)

- Rindflesch, Thomas C., and Alan R. Aronson. 1993. Semantic processing in information retrieval. Charles Safran (ed.) *Proceedings of the 17th Annual Symposium on Computer Applications in Medical Care*, 611-15.
- Aronson, Alan R.; Thomas C. Rindflesch; and Allen C. Browne. 1994. Exploiting a large thesaurus for information retrieval. *Proceedings of RIAO*, 197-216.
- Rindflesch, Thomas C., and Alan R. Aronson. 1994. Ambiguity resolution while mapping free text to the UMLS Metathesaurus. Judy G. Ozbolt (ed.) *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, 240-4.
- Rindflesch, Thomas C. 1995. Integrating natural language processing and biomedical domain knowledge for increased information retrieval effectiveness. *Proceedings of the 5th Annual Dual- use Technologies and Applications Conference*, 260-5.

- Rindflesch, Thomas C. 1996. Natural language processing. William Grabe (ed.) *Annual Review of Applied Linguistics* 16:71-85. Cambridge: Cambridge University Press.
- Sneiderman, Charles A.; Thomas C. Rindflesch; and Alan R. Aronson. 1996. Finding the findings: Identification of findings in medical literature using restricted natural language processing. James J. Cimino (ed.) *Proceedings of the AMIA Annual Fall Symposium*, 239-43.
- Aronson, Alan R., and Thomas C. Rindflesch. 1997. Query expansion using the UMLS Metathesaurus. Daniel R. Masys (ed.) *Proceedings of the AMIA Annual Fall Symposium*, 485-9.
- Bean, Carol A.; Thomas C. Rindflesch; and Charles A. Sneiderman. 1998. Automatic semantic interpretation of anatomical spatial relationships in clinical text. Christopher G. Chute (ed.) *Proceedings of the AMIA Annual Symposium*, 897-901.
- Divita, Guy; Allen C. Browne; and Thomas C. Rindflesch. 1998. Evaluating lexical variant generation to improve information retrieval. Christopher G. Chute (ed.) *Proceedings of the AMIA Annual Symposium*, 775-9.
- Sneiderman, Charles A.; Thomas C. Rindflesch; and Carol A. Bean. 1998. Identification of anatomical terminology in medical text. Christopher G. Chute (ed.) *Proceedings of the AMIA Annual Symposium*, 428-32.
- Wright, Lawrence W.; Holly K. Grosetta Nardini; Alan R. Aronson; and Thomas C. Rindflesch. 1999. Hierarchical concept indexing of full-text documents in the Unified Medical Language System Information Sources Map. *Journal of the American Society for Information Science* 50(6):514-23.
- Rindflesch, Thomas C.; Lawrence Hunter; and Alan R. Aronson. 1999. Mining molecular binding terminology from biomedical text. Nancy M. Lorenzi (ed.) *Proceedings of the AMIA Annual Symposium*, 127-31.
- Rindflesch, Thomas C.; Lorraine Tanabe; John N. Weinstein; and Lawrence Hunter. 2000. EDGAR: Extraction of drugs, genes, and relations from the biomedical literature. *Pacific Symposium on Biocomputing (PSB)* 5:514-25.
- Rindflesch, Thomas C.; Jayant V. Rajan; and Lawrence Hunter. 2000. Extracting molecular binding relationships from biomedical text. *Proceedings of the 6th Applied Natural Language Processing Conference*, 188-95. Association for Computational Linguistics.
- Rindflesch, Thomas C.; Carol A. Bean; and Charles A. Sneiderman. 2000. Argument identification for arterial branching predication asserted in cardiac catheterization reports. J. Marc Overhage (ed.) *Proceedings of the AMIA Annual Symposium*, 704-8.
- Aronson, Alan R.; Olivier Bodenreider; H. Florence Chang; Susanne M. Humphrey; James G. Mork; Stuart J. Nelson; Thomas C. Rindflesch; and W. John Wilbur. 2000. The NLM indexing initiative. J. Marc Overhage (ed.) *Proceedings of the AMIA Annual Symposium*, 17-21.
- Humphrey, Susanne M.; Thomas C. Rindflesch; and Alan R. Aronson. 2000. Automatic indexing by discipline and high-level category: Methodology and potential applications. *Proceedings of the 11th SIG/CR Classification Research Workshop*, 103-16.

- Bodenreider, Olivier; Anita Burgun; and Thomas C. Rindflesch. 2001. Lexically-suggested hyponymic relations among medical terms and their representation in the UMLS. Proceedings of Terminology and Artificial Intelligence Conference, 11-21.
- Bodenreider, Olivier; Anita Burgun; and Thomas C. Rindflesch. 2002. Assessing the consistency of a biomedical terminology through lexical knowledge. Proceedings of the Workshop on Natural Language Processing in Biomedical Applications. European Federation for Medical Informatics.
- Bodenreider, Olivier; Thomas C. Rindflesch; and Anita Burgun. 2002. Unsupervised corpus-based method for extending a biomedical terminology. Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, 53-60. Association for Computational Linguistics.
- Libbus, Bisharah, and Thomas C. Rindflesch. 2002. NLP-based information extraction for managing the molecular biology literature. Isaac Kohane (ed.) Proceedings of the AMIA Annual Symposium, 445-9.
- Sarkar, Indra Neil, and Thomas C. Rindflesch. 2002. Discovering protein similarity using natural language processing. Isaac Kohane (ed.) Proceedings of the AMIA Annual Symposium, 677-81.
- Srinivasan, Padmini, and Thomas C. Rindflesch. 2002. Exploring text mining from MEDLINE. Isaac Kohane (ed.) Proceedings of the AMIA Annual Symposium, 722-6.
- Srinivasan, Suresh; Thomas C. Rindflesch; William T. Hole; and Alan R. Aronson. 2002. Finding UMLS Metathesaurus concepts in MEDLINE. Isaac Kohane (ed.) Proceedings of the AMIA Annual Symposium, 727-31.
- Rindflesch, Thomas C., and Alan R. Aronson. 2002. Semantic processing for enhanced access to biomedical knowledge. Vipul Kashyap and Leon Shklar (eds.) Real World Semantic Web Applications, 157-72. IOS Press.
- Fiszman, Marcelo; Thomas C. Rindflesch; and Halil Kilicoglu. 2003. Integrating a hypernymic proposition interpreter into a semantic processor for biomedical text. Proceedings of the AMIA Annual Symposium, 239-43.
- Rindflesch, Thomas C.; Bisharah Libbus; Dimitar Hristovski; Alan R. Aronson; and Halil Kilicoglu. 2003. Semantic relations asserting the etiology of genetic diseases. Proceedings of the AMIA Annual Symposium, 554-8.
- Rindflesch, Thomas C., and Marcelo Fiszman. 2003. The interaction of domain knowledge and linguistic structure in natural language processing: Interpreting hypernymic propositions in biomedical text. Journal of Biomedical Informatics 36(6):462-77.
- Chapman, Wendy W.; Marcelo Fiszman; John N. Dowling; Brian E. Chapman; and Thomas C. Rindflesch. 2004. Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap. Medinfo [in press].
- Fiszman, Marcelo; Thomas C. Rindflesch; and Halil Kilicoglu. 2004. Summarization of an online medical encyclopedia. Medinfo [in press].
- Leroy, Gondy, and Thomas C. Rindflesch. 2004. Using symbolic knowledge in the UMLS to disambiguate words in small datasets with a naive Bayes classifier. Medinfo [in press].

- Smith, Lawrence H.; Thomas C. Rindflesch; and W. John Wilbur. 2004. MedPost: A part of speech tagger for biomedical text. Bioinformatics [in press].
- Fiszman, Marcelo; Thomas C. Rindflesch; and Halil Kilicoglu. 2004. Abstraction summarization for managing the biomedical research literature. HLT/NAACL Workshop on Computational Lexical Semantics [in press].
- Libbus, Bisharah; Halil Kilicoglu; Thomas C. Rindflesch; James G. Mork; and Alan R. Aronson. 2004. Using natural language processing, Locus Link and the Gene Ontology to compare OMIM to MEDLINE. HLT/NAACL Workshop on Linking the Biological Literature, Ontologies and Databases: Tools for Users [in press].